Belle II Computing Model

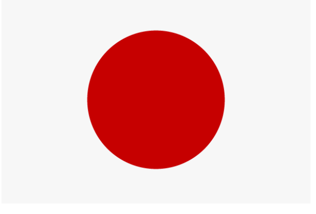
Thomas Kuhr

Computing Workshop

18.06.2010

World Cup

Japan : Cameroon



1 : 0

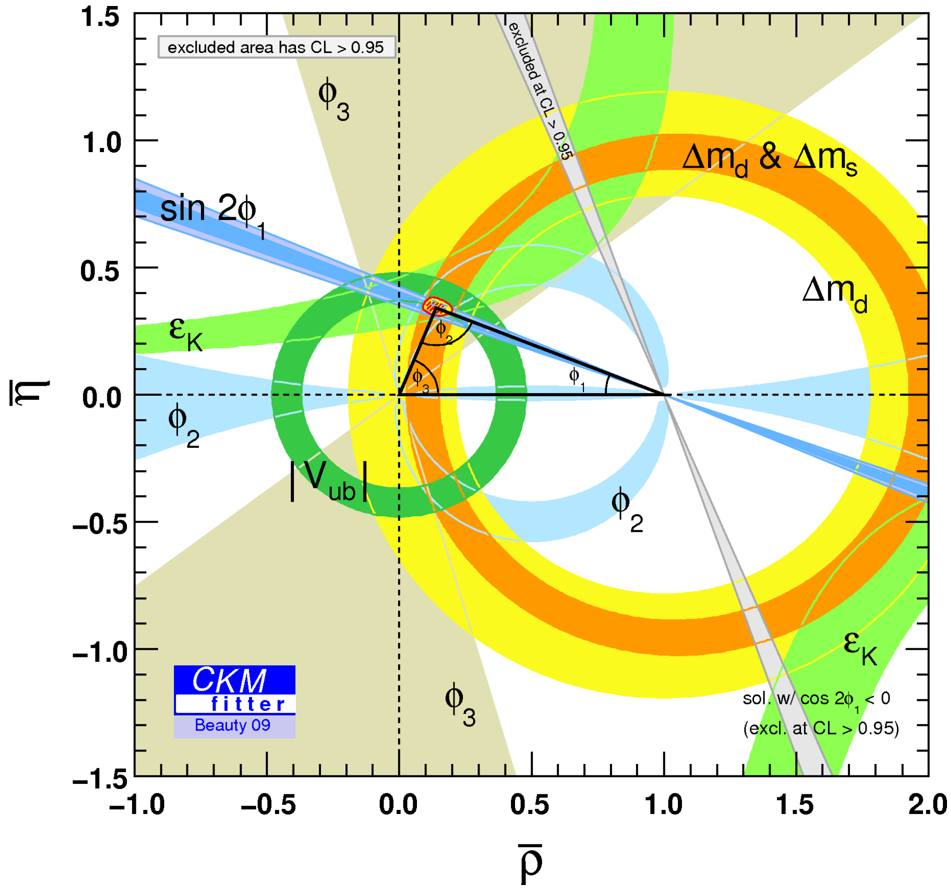
Italy : Paraguay

1 : 1

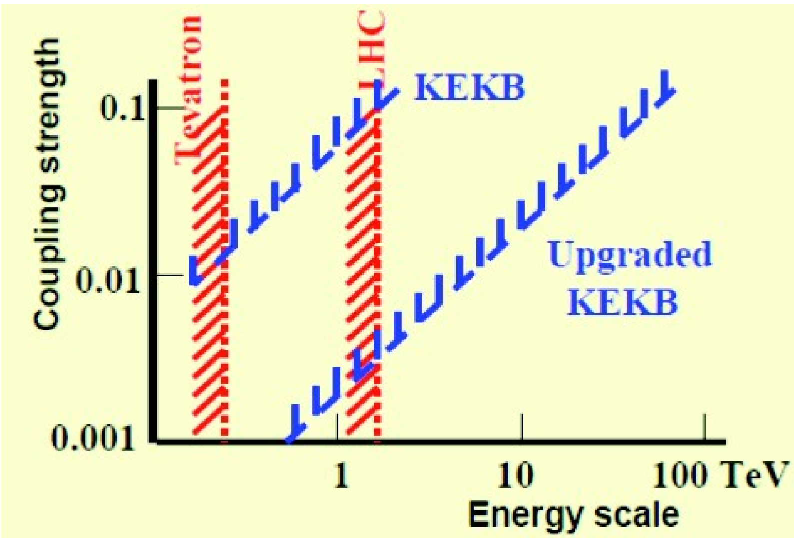
Physics Objective of Belle and Belle II

- ✔ Confirmation of KM mechanism of ~~CP~~ in the Standard Model
- ✗ CP violation in the SM by many orders of magnitude too small to generate observed baryon asymmetry in the universe
- ➢ Need sources of CP violation beyond the SM

➔ Super B factory



Flavor Physics in the LHC Era



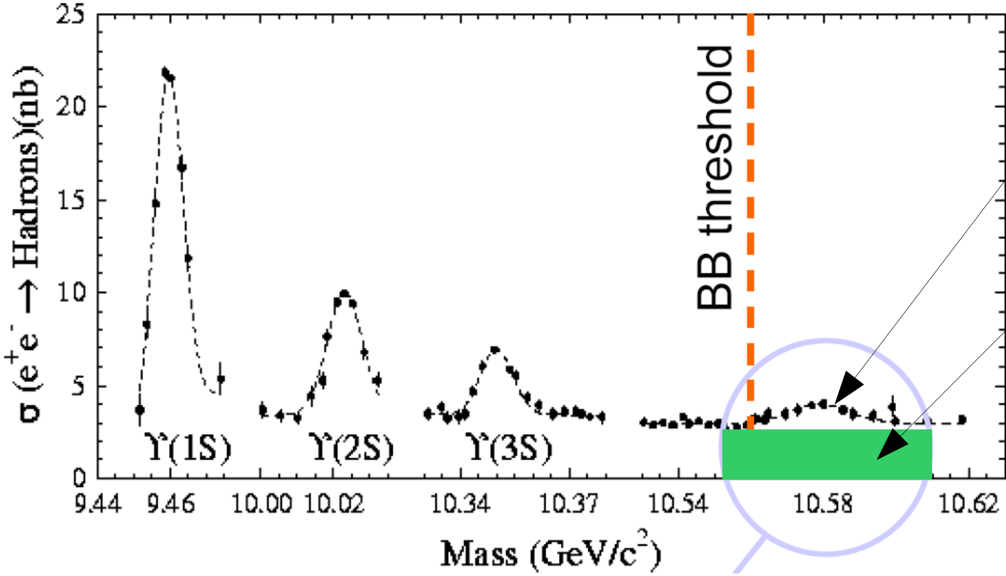
- Complementary to direct search for NP at LHC

→ "DNA test" of NP

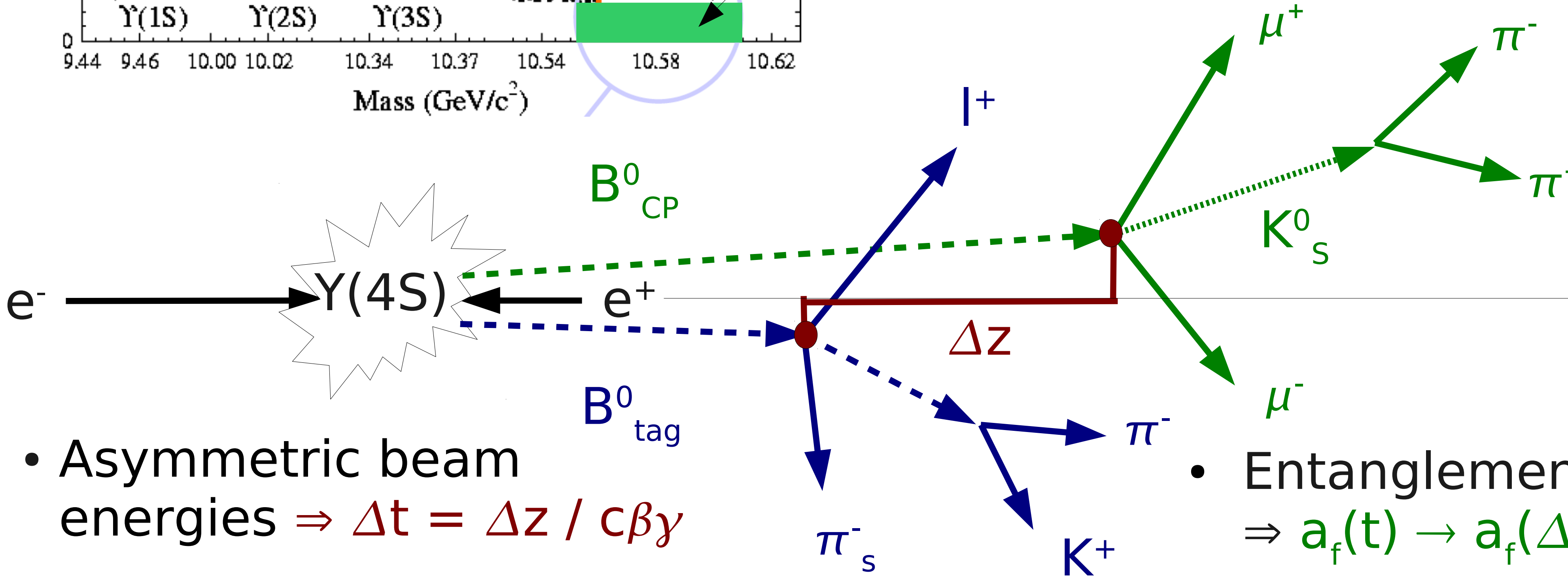
	AC	RVV2	AKM	δLL	FBMSSM	LHT	RS
$D^0-\bar{D}^0$	★★★	★	★	★	★	★★★	?
ϵ_K	★	★★★	★★★	★	★	★★	★★★
$S_{\psi\phi}$	★★★	★★★	★★★	★	★	★★★	★★★
$S_{\phi K_S}$	★★★	★★	★	★★★	★★★	★	?
$A_{\rm CP}(B\to X_s\gamma)$	★	★	★	★★★	★★★	★	?
$A_{7,8}(B\to K^*\mu^+\mu^-)$	★	★	★	★★★	★★★	★★	?
$A_9(B\to K^*\mu^+\mu^-)$	★	★	★	★	★	★	?
$B\to K^{(*)}\nu\bar{\nu}$	★	★	★	★	★	★	★
$B_s\to\mu^+\mu^-$	★★★	★★★	★★★	★★★	★★★	★	★
$K^+\to\pi^+\nu\bar{\nu}$	★	★	★	★	★	★★★	★★★
$K_L\to\pi^0\nu\bar{\nu}$	★	★	★	★	★	★★★	★★★
$\mu\to e\gamma$	★★★	★★★	★★★	★★★	★★★	★★★	★★★
$\tau\to\mu\gamma$	★★★	★★★	★	★★★	★★★	★★★	★★★
$\mu+N\to e+N$	★★★	★★★	★★★	★★★	★★★	★★★	★★★
d_n	★★★	★★★	★★★	★★	★★★	★	★★★
d_e	★★★	★★★	★★	★	★★★	★	★★★
$(g-2)_\mu$	★★★	★★★	★★	★★★	★★★	★	?

Altmannshofer, Buras, Gori, Paradis and Straub
Nucl.Phys.B830, 17-94, 2010

Time Dependent ~~CP~~ Measurement

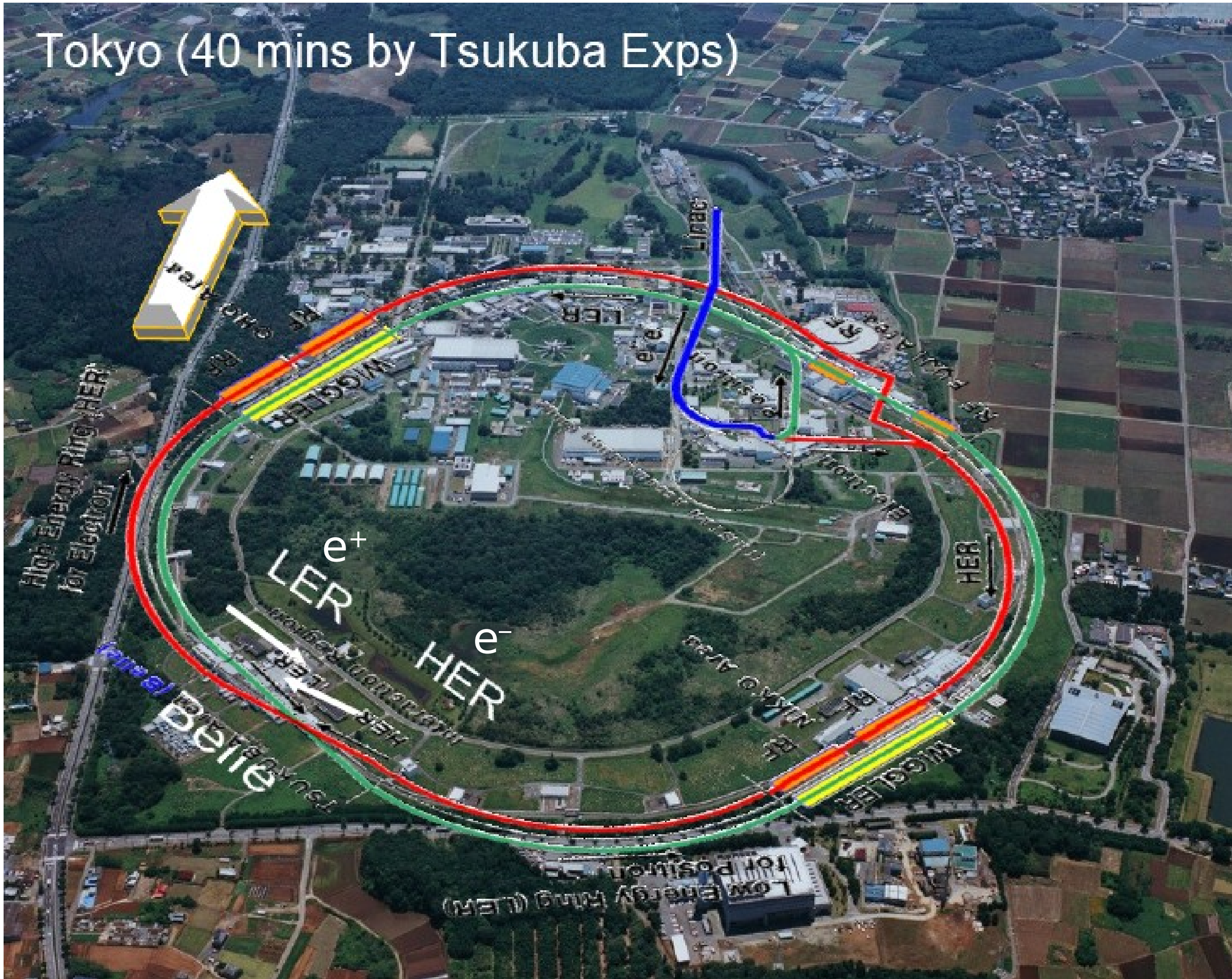


- $e^+e^- \rightarrow Y(4S) \rightarrow B^0\bar{B}^0$ (50%)
  $\rightarrow B^+B^-$ (50%)
- Continuum background

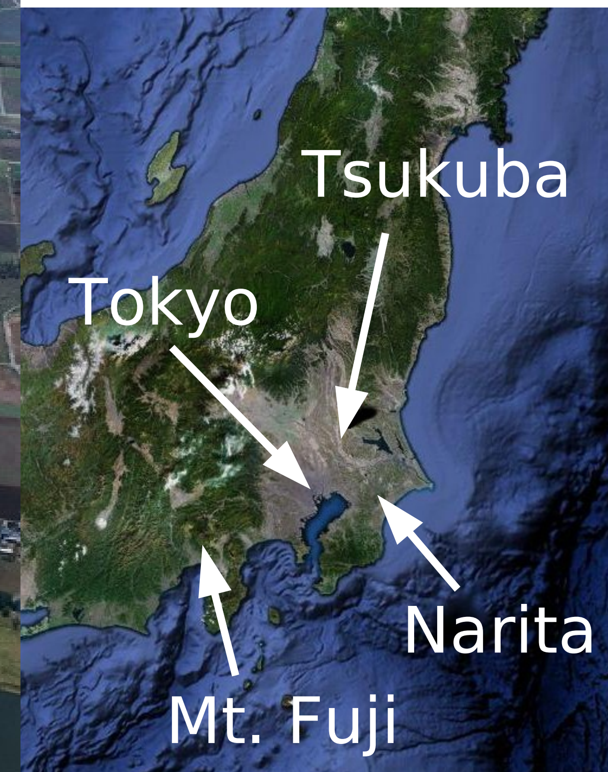


- Asymmetric beam energies $\Rightarrow \Delta t = \Delta z / c\beta\gamma$
- Entanglement $\Rightarrow a_f(t) \rightarrow a_f(\Delta t)$

KEK Site

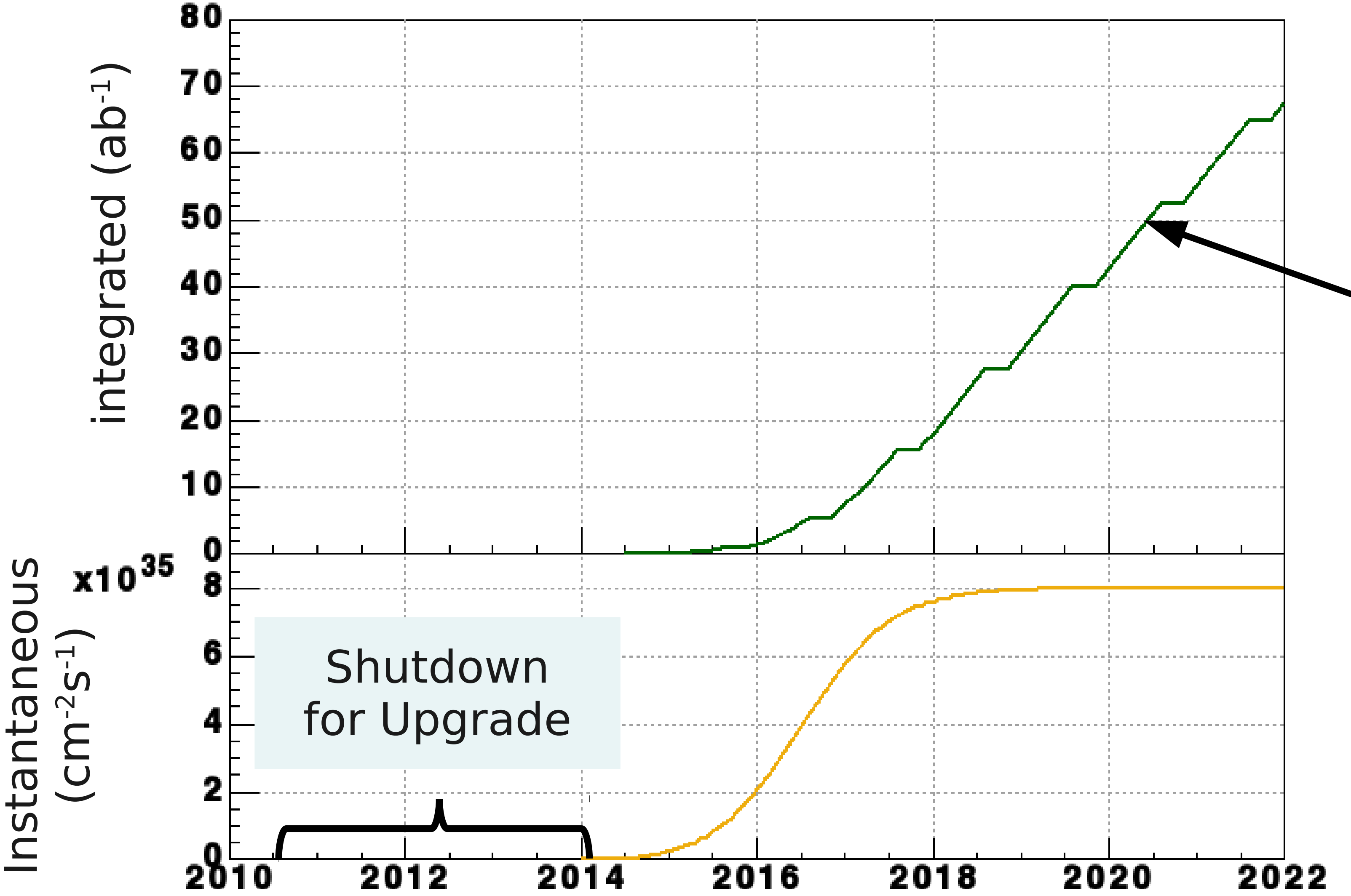


Tokyo (40 mins by Tsukuba Exps)
e+
LER
e−
HER
Belle
Linac
WIGGLER
RF
HER



Tsukuba
Tokyo
Narita
Mt. Fuji

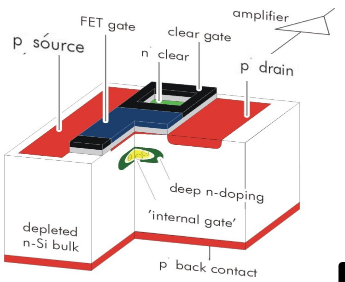
Projection of Luminosity

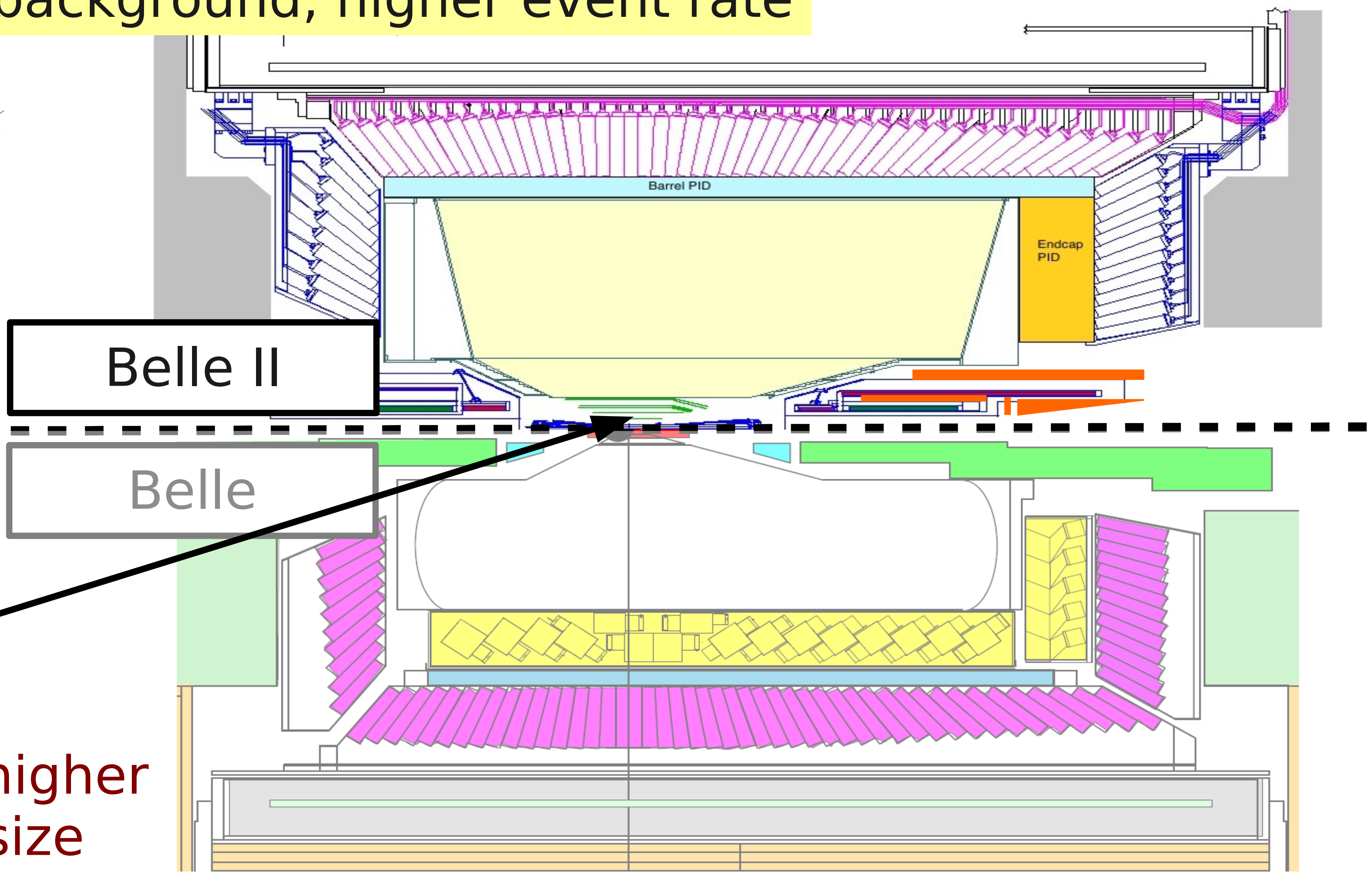


50 times Belle data

Belle II Detector

➔ Higher background, higher event rate





New
2 layer
DEPFET
pixel
detector

→ much higher
event size

Belle II Collaboration

- June 2004: Letter of Intent
- March 2008: First proto collaboration meeting
- December 2008: New collaboration founded

- ~300 members
  47 institutes from 13 countries

Belle II

International Collaboration

Project Status

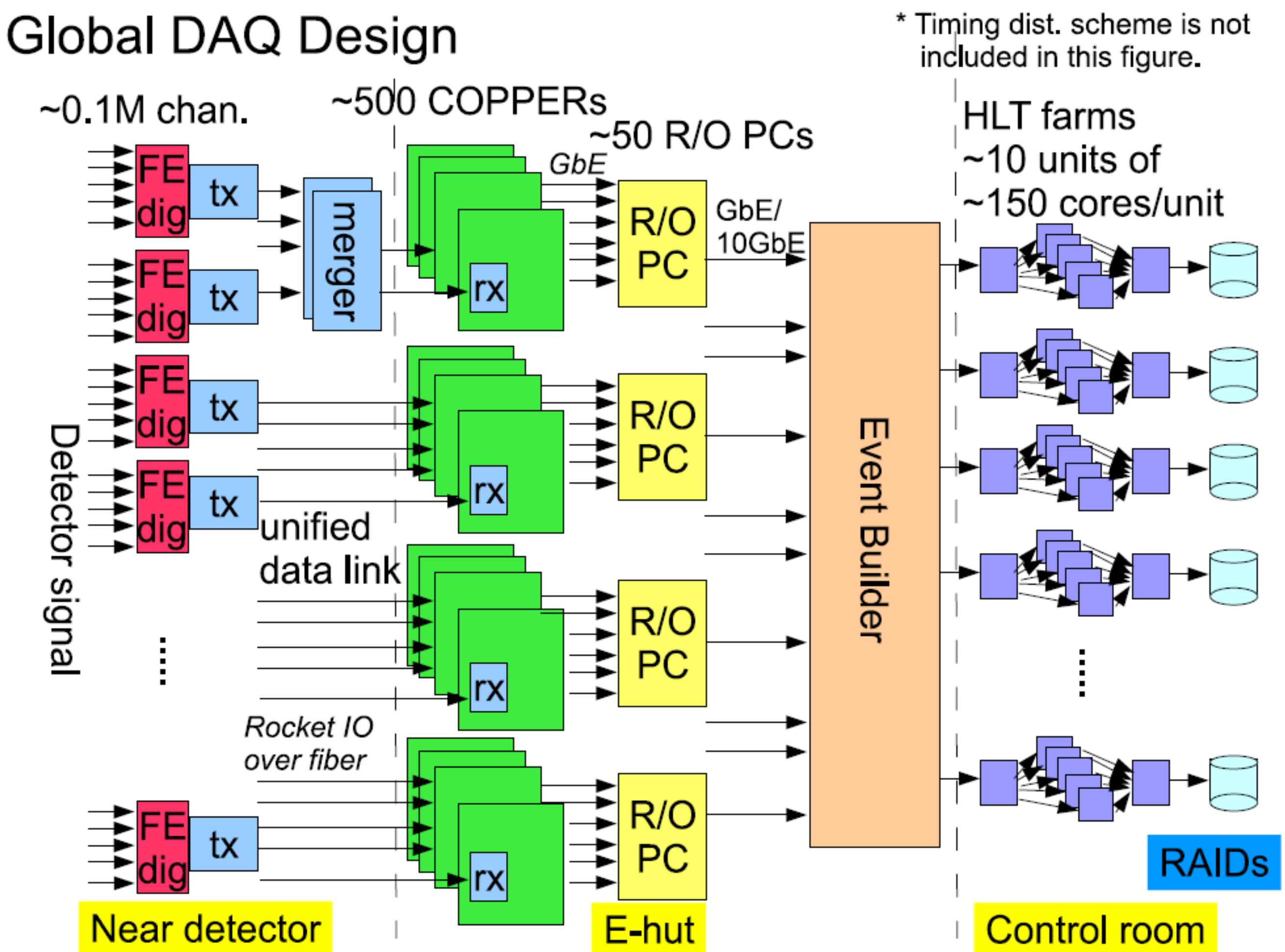
- ✔ Preliminary approval by Japanese government in January
- Final decision expected soon
- ✔ TDR is written (~480 pages)
  - ➔ Reviewed three weeks ago
  - ➔ To be published in July
- Belle data taking ends in June
- ➔ Then upgrade work can start
- ➢ Well on track to start data taking in 2014


Belle II

Technical Design Report

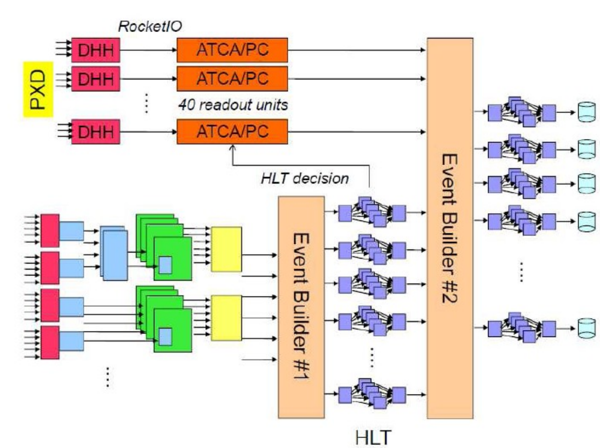
Version: 1.1
May 7, 2010

Belle II Data Acquisition / High Level Trigger



Challenges:

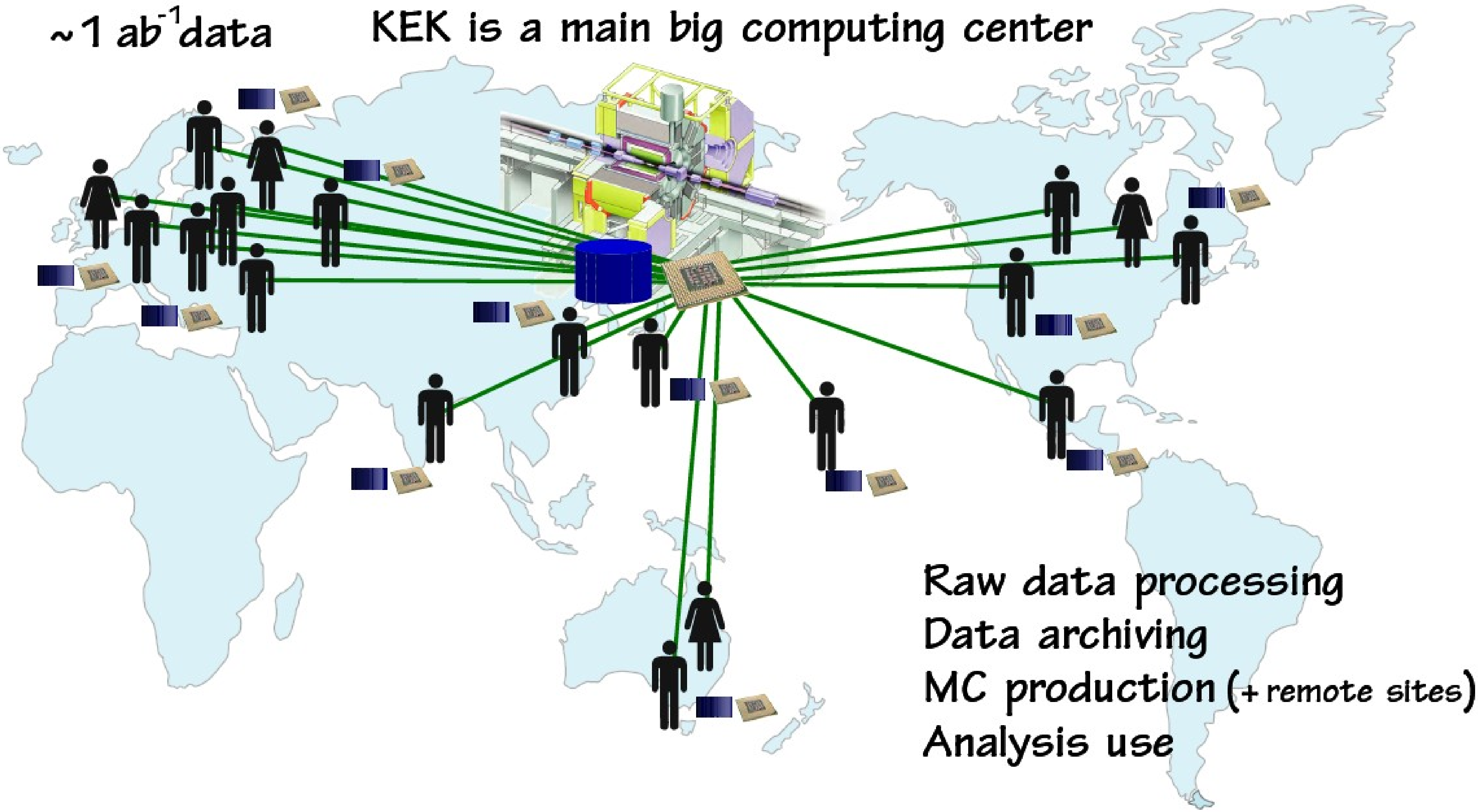
- high input rate up to 30 kHz
- PXD event data reduction
  1 MB → ~200 kB
  (all other Belle II detectors together ~100 kB)



Estimated Data Rates

Experiment	Rate [Hz]	Event Size [kB]	Rate [MB/s]
High rate scenario for Belle II DAQ			
Belle II	6,000	300	1,800
LCG TDR (2005)			
ALICE (HI)	100	12,500	1,250
ALICE (pp)	100	1,000	100
ATLAS	200	1,600	320
CMS	150	1,500	225
LHCb	2,000	25	50

Belle Computing: Centralized at KEK



Considerations for Belle II Computing

- 50 times more data in ~10 years
  - ➔ O(50) times more computing resources
- Can not expect KEK to provide all resources
- *Have to enable all Belle II member institutes to contribute*
  - ➔ **Distributed computing system** based on grid services (gLite)
  - ➢ Can benefit from existing LCG infrastructure
  - ➢ Profit from experience of LHC experiments and their well-established and matured solutions
- Tasks: Data processing, MC production, physics analysis

Raw Data Processing

- Huge data size (raw data rate up to 1.8 GB/s)
  → comparable to LHC experiments
- Write once, read only few times in managed way

➔ Tape as storage medium

  ➢ Most cost and power efficient,
    but high maintenance effort

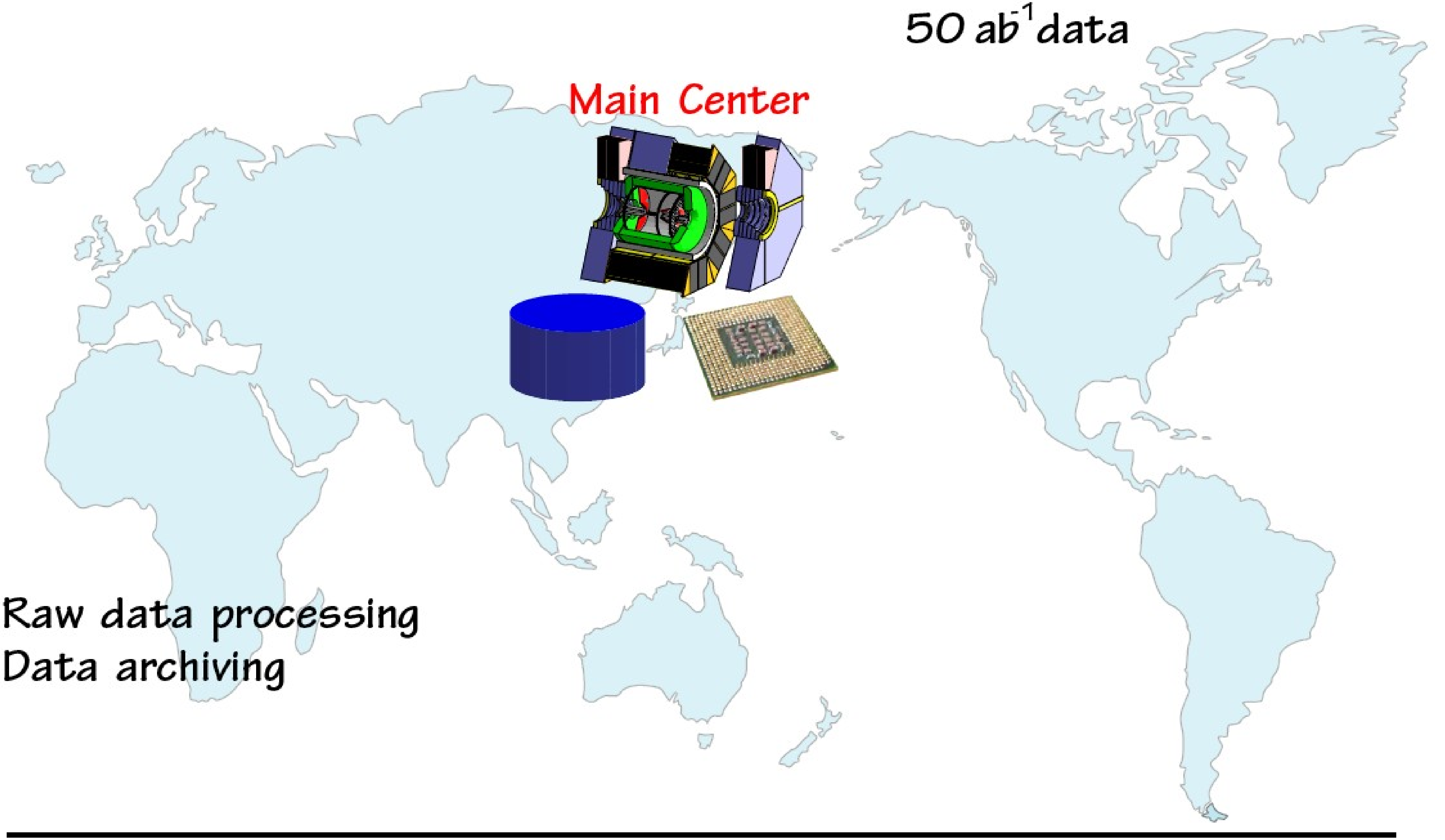
- Source of raw data is Belle II detector

➔ Store and process at KEK, no replication to remote sites

  ➢ Would require huge network bandwidth, high operation efficiency, maintenance of tape system

➢ Simpler than LCG model

Raw Data Processing



MC Production

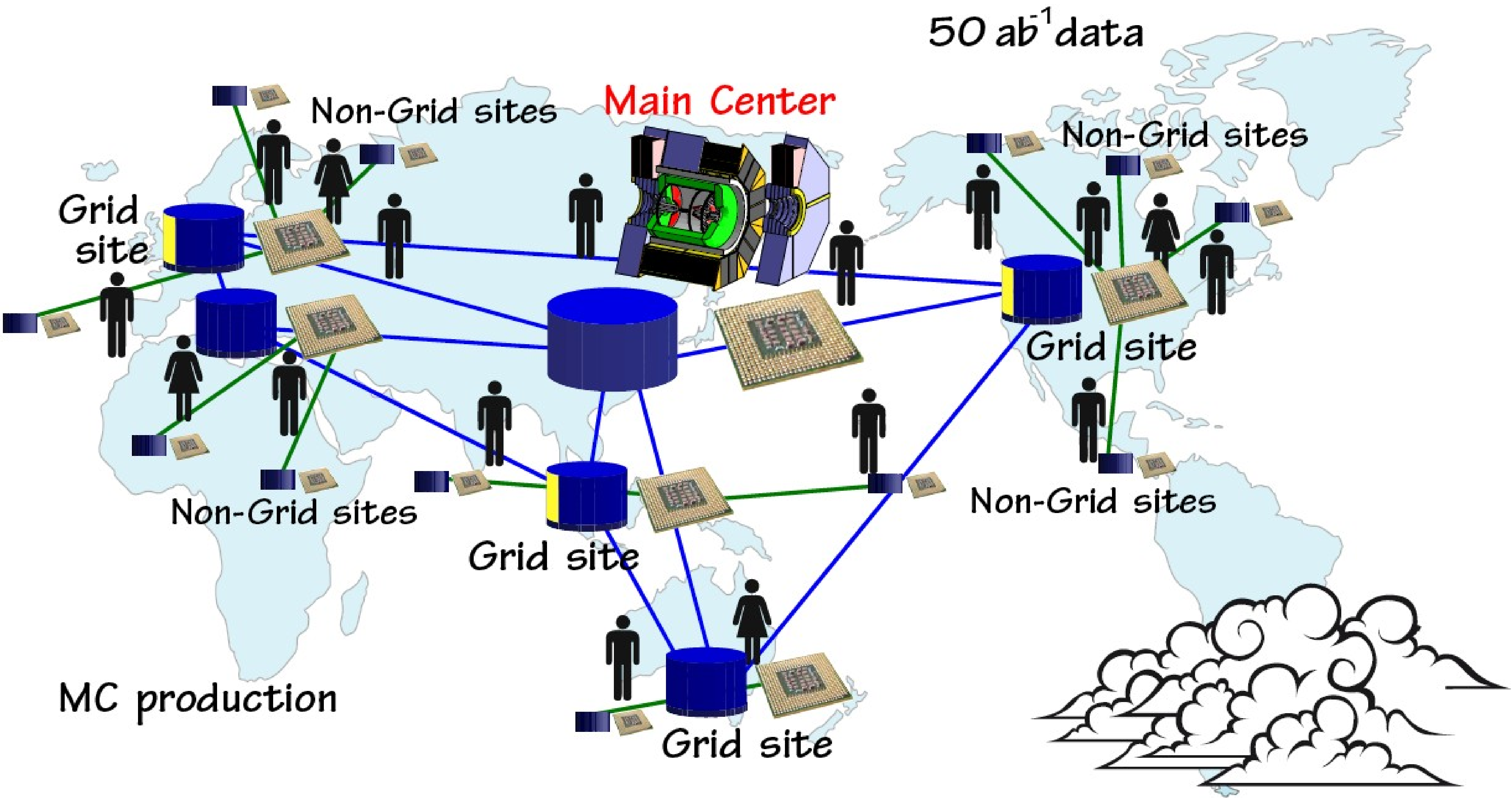
- Run-dependent, generic MC:
  $B^0 B^0$, $B^+ B^-$, charm, and uds events
- Corresponding to *N* times the real data size
- Produced in managed way
- No input data needed
  (except generator and background data files)

➔ Well suited for a distributed environment

  - Including cloud computing resources

➔ Output stored at grid sites where it is produced

MC Production



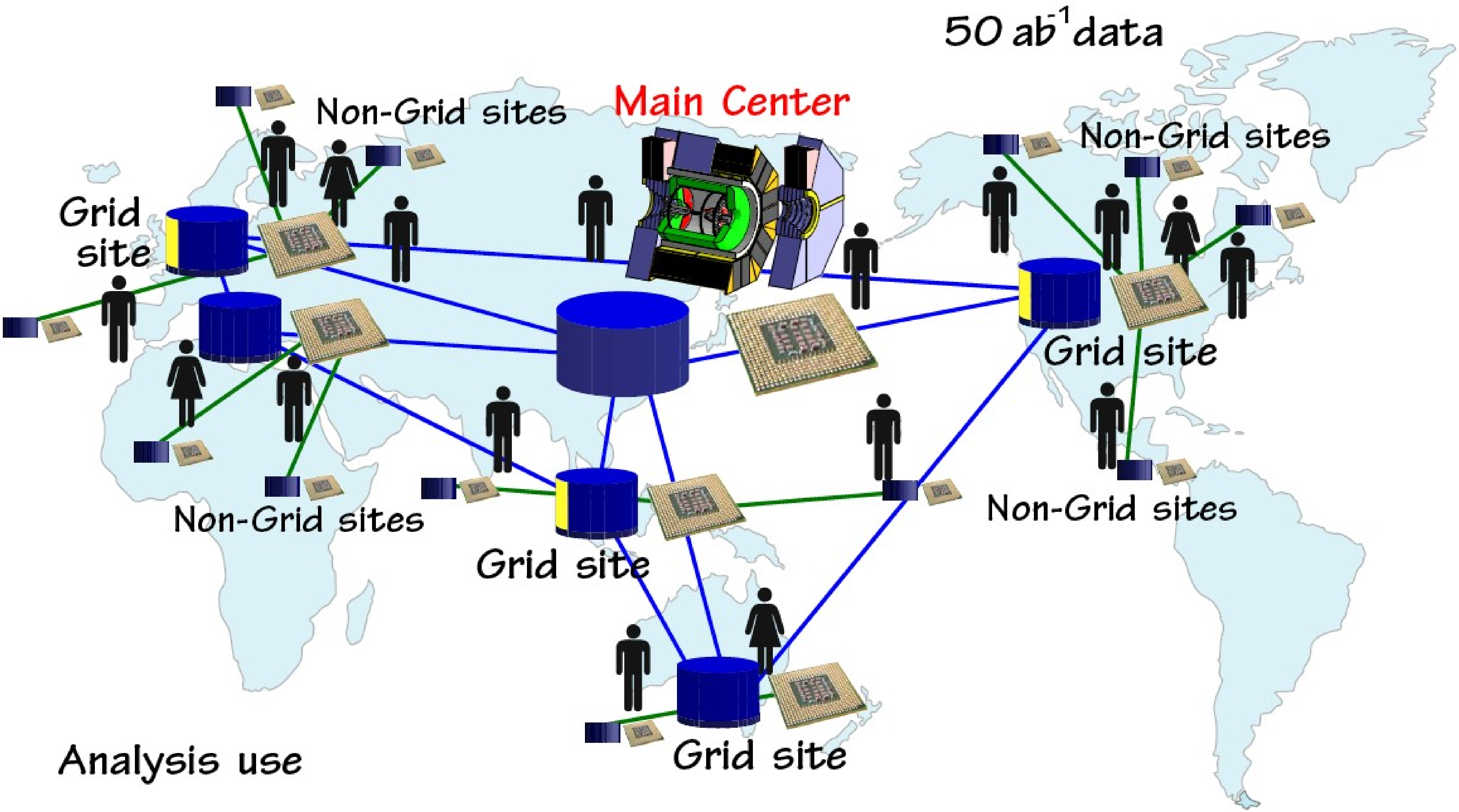
Physics Analysis

- Need real / MC data in mDST format as input
- ➔ Copy mDST real data to grid sites
- ➔ Replicate MC datasets in a managed way if needed
  - ➢ Produce ntuples on the grid

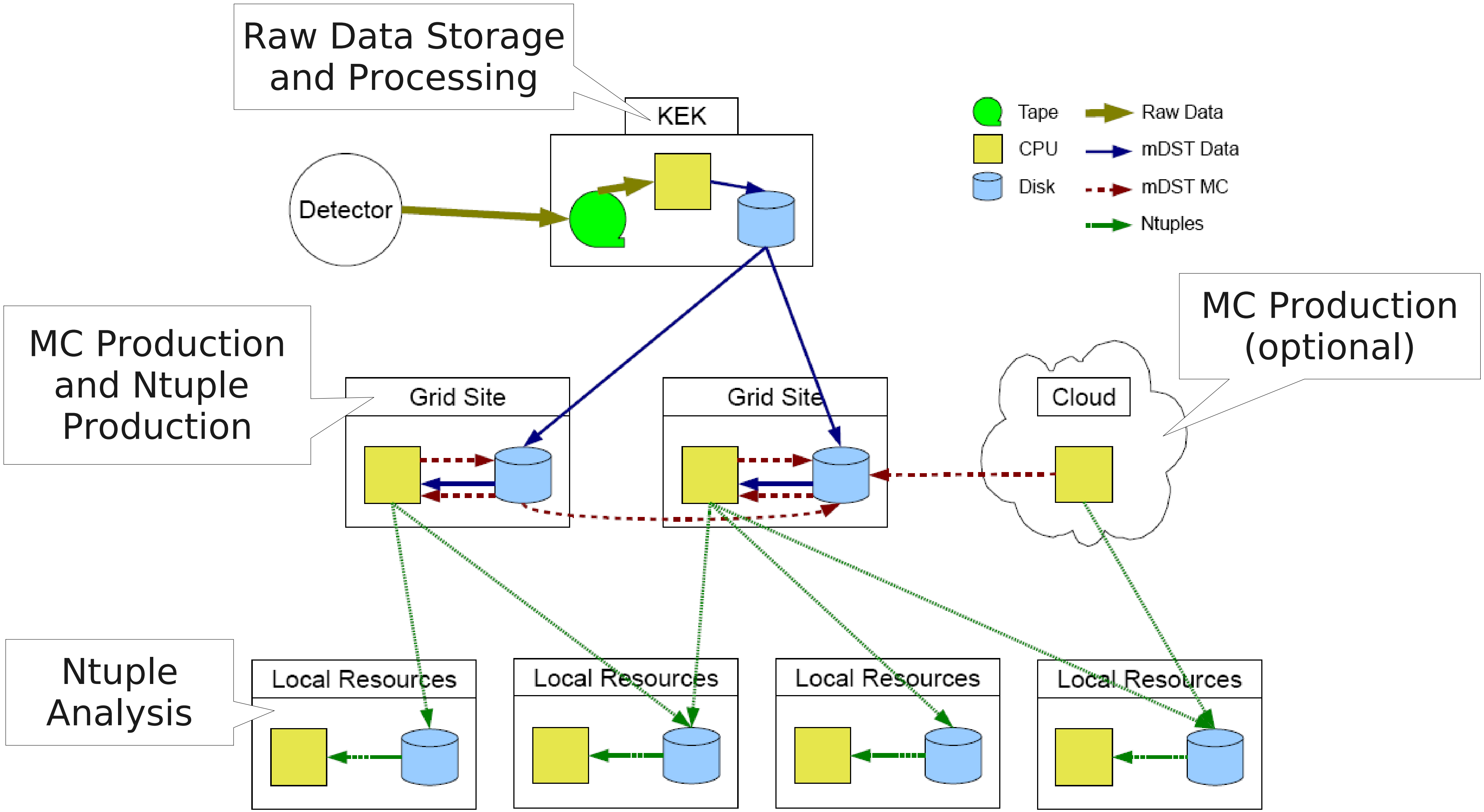
- Random, uncoordinated access
- ➔ Store mDST real / MC data on disk

- Need fast turn-around time for ntuple analysis
- ➔ Copy them to local resources
  - ➢ Do ntuple analysis on local resources

Physics Analysis



Belle II Computing Model



Tasks of Computing Facilities

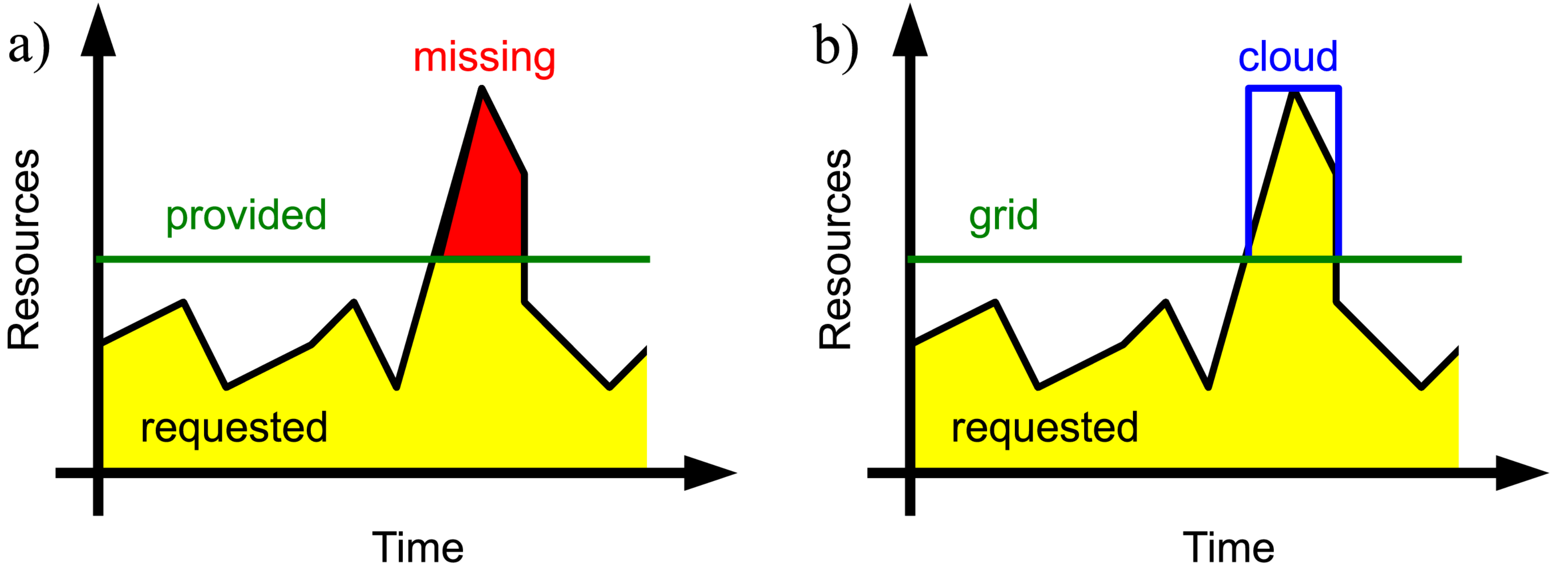
Non-grid Sites	Grid Sites	KEK	
		Storage and Processing of Raw Data	Main Center
	Experiment-specific Services	Experiment-specific Services	
	Monte-Carlo Production	Monte-Carlo Production	Grid
	Data Analysis	Data Analysis	
Ntuple-level Analysis	Ntuple-level Analysis	Ntuple-level Analysis	Local Resources
User Interface	User Interface	User Interface	

(Commercial) Cloud Computing

- Resource demands vary with time
- Fair-share can solve this issue only to some extent

➔ Cloud computing allows to buy resources on demand

  - Well suited to absorb peaks in varying resource demand



Cloud Computing in Belle II

- Risk: vendor lock-in
  - ➔ No permanent data storage on the cloud
  - ➔ Much less critical for CPU resources
- Large data transfer / storage not cost efficient (now)
  - ➔ Use cloud primarily for MC production
  - ➔ No data processing
  - ➔ Maybe physics analysis
- Accounting issues

➢ Baseline of computing resources provided by the grid

➢ Cloud computing is option for peak demands

Status of Cloud Computing Project @ IFJ PAN

- Timeline of the system development:

  Phase1 (→ 1Q 2011): - evaluation of OpenSource Cloud Managers
  - evaluation of virtualization engines,
  - development of tools and interfaces,
  - establish functional system for local users

  Phase2 (→ 1Q 2012): distributed Cloud

  Phase3 (→ 1Q 2013): - production system (also for Belle II use)
  - admitt IT startups

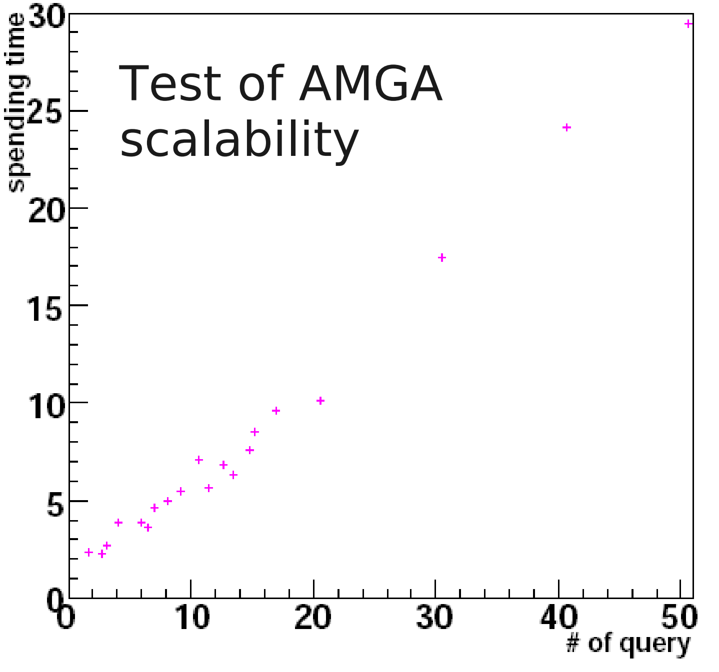
- Phase1 till now: - 3 testbeds, 2 blades each (16 cores)
  - evaluation teams (1/testbed):
  Eucalyptus, OpenNebula, Nimbus

Metadata Catalog

- Task:
  Identify files matching physics selection criteria (dataset)

➔ AMGA (Arda Metadata Grid Application)

  - Official gLite metadata service
  - Support for several
    DB backends
  - Grid-based authentication
  - Replication



➢ File level metadata: O(60 GB)

➢ Event level metadata investigated: O(20 TB)

Belle LSDH & NSDC

Belle Data Transfer	• KEK → KISTI • Currently, 147TB data are archived	• 93% of planned done
Metadata Extraction		
Scalability Test		
Metadata Replication		
Grid-awaring of Belle data	• SRMDoor & Disk Pool (SEDPM) • File registration to SE without copy • Registration to LFC • Co-test with SE & LFC	• SRM Door & Disk Pool → Done • File Registration to SE → Done • Registration to LFC → Done for KEK & NSDC • Test with SE & LFC → Simple test done • Outside test → done
Test on Grid Env.	• Job submission to Grid (with access to AMGA, SE, LFC)	• Basic Test was successful • Every Testers should have grid cert. and Belle VO membership
What Next ?	• Any other Belle2 SW tests on Grid?	

Data handling – features

- Parallel replication of files and datasets
- Listing datasets
- LFN → GUID conversion
- GUID → LFN conversion
- Listing replicas of file
- Uploading files and datasets to SE
- Deleting files and datasets

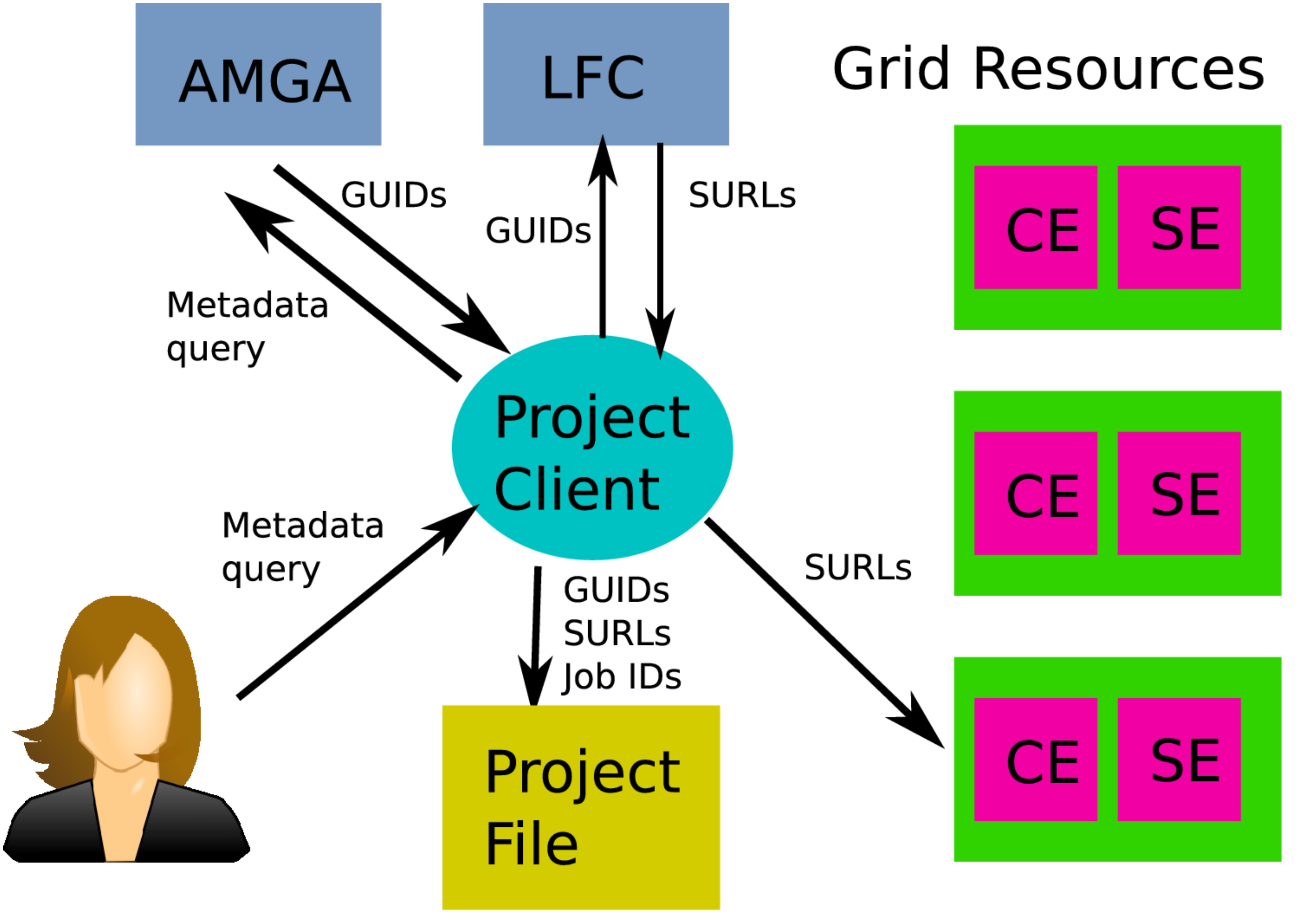
administrator

internal use

Analysis Model

- Task: Process selected set of files with given analysis code
  - ➔ Project
- ➢ Have to make sure each file is processed exactly once
- No solution on project level provided by grid services
- Start with implementation of simple solution
  - Static job – data assignment
  - Job push model
- Then gradually add more complexity
  - Dynamic job – data assignment (a la SAM, CDF/D0)
  - Job pull model (pilot agents)

Project Client



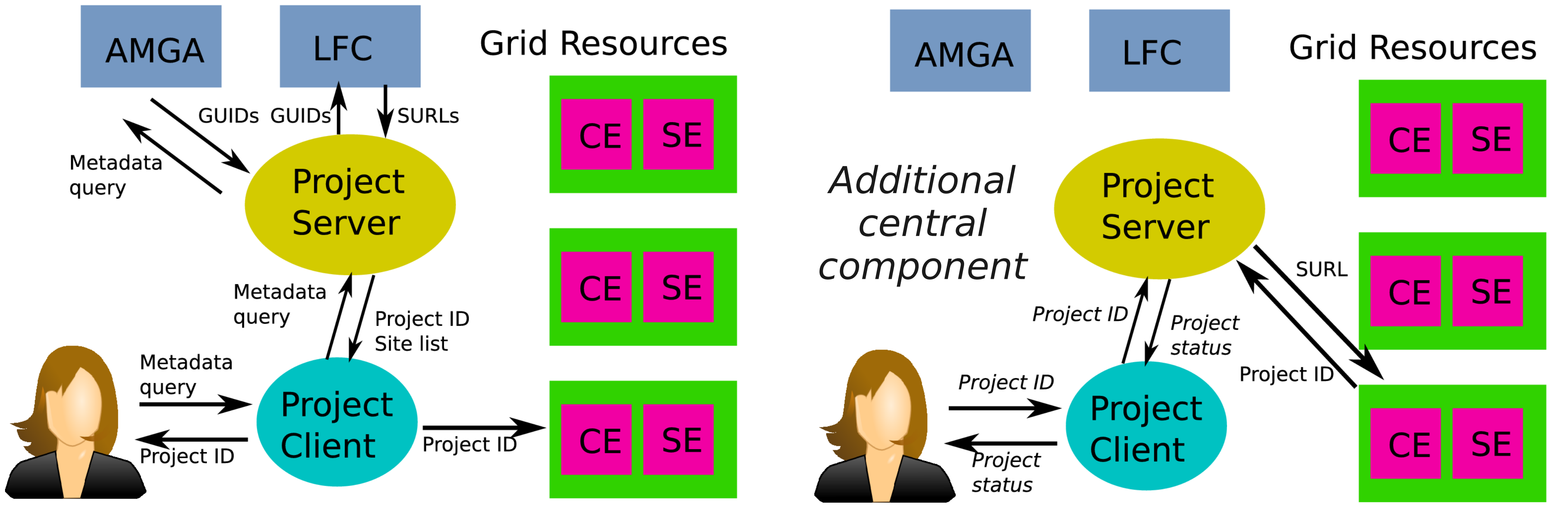
Bookkeeping using AMGA

- Why?
  - We know it
    - Configuration
    - API
  - It has built in-replication
- Why not?
  - It's mainly for metadata
  - No full synchronization, only master-slave replication
  - Maybe PostreSQL could be better?
- What we need to do?
  - Check if it fulfils our needs
  - Test it!

Job submission steps

- (Cream) Proxy delegation on specified CE (if we don't already have one)
- Register job to CE
- Upload files to InputSandBox
- Start the job
- Monitor the job status
- Download the job output

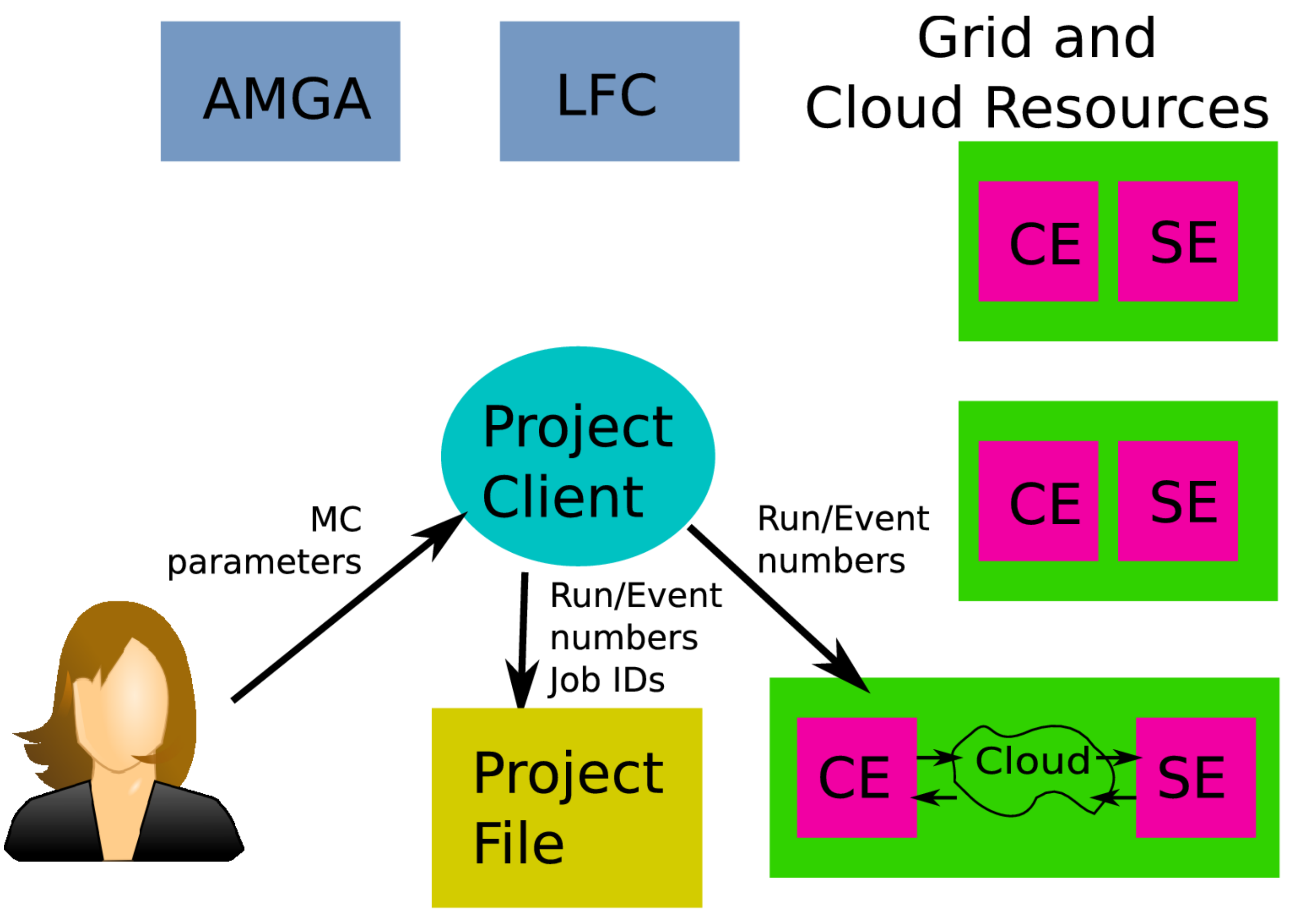
Project Server: Dynamic Data Assignment



- ✔ Easier job submission (identical jobs)
- ✔ More fine grained monitoring (processed files)
- ✔ Only running jobs at working sites get data assigned
- ✔ Faster nodes get more input files
  → Automatic load balancing, reduced overall time to finish

Same Strategy for MC Production

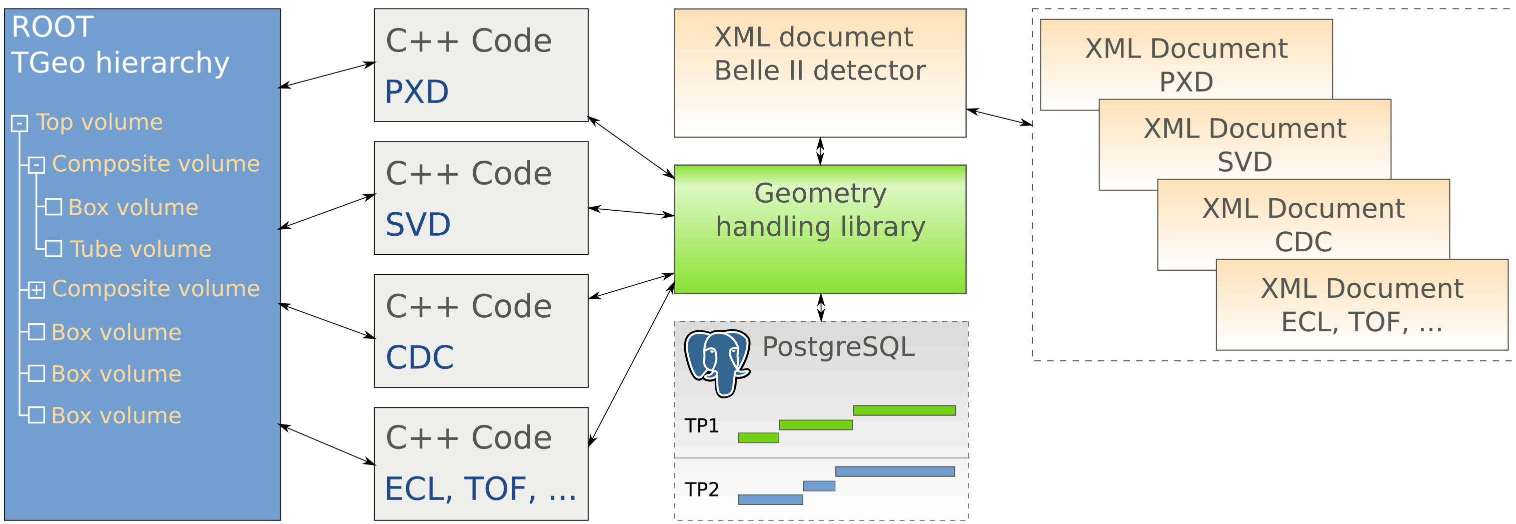
- Input data file is replaced by run number, #events



Geometry and Conditions Database

- (Time-dep.) parameters stored in XML files and DB

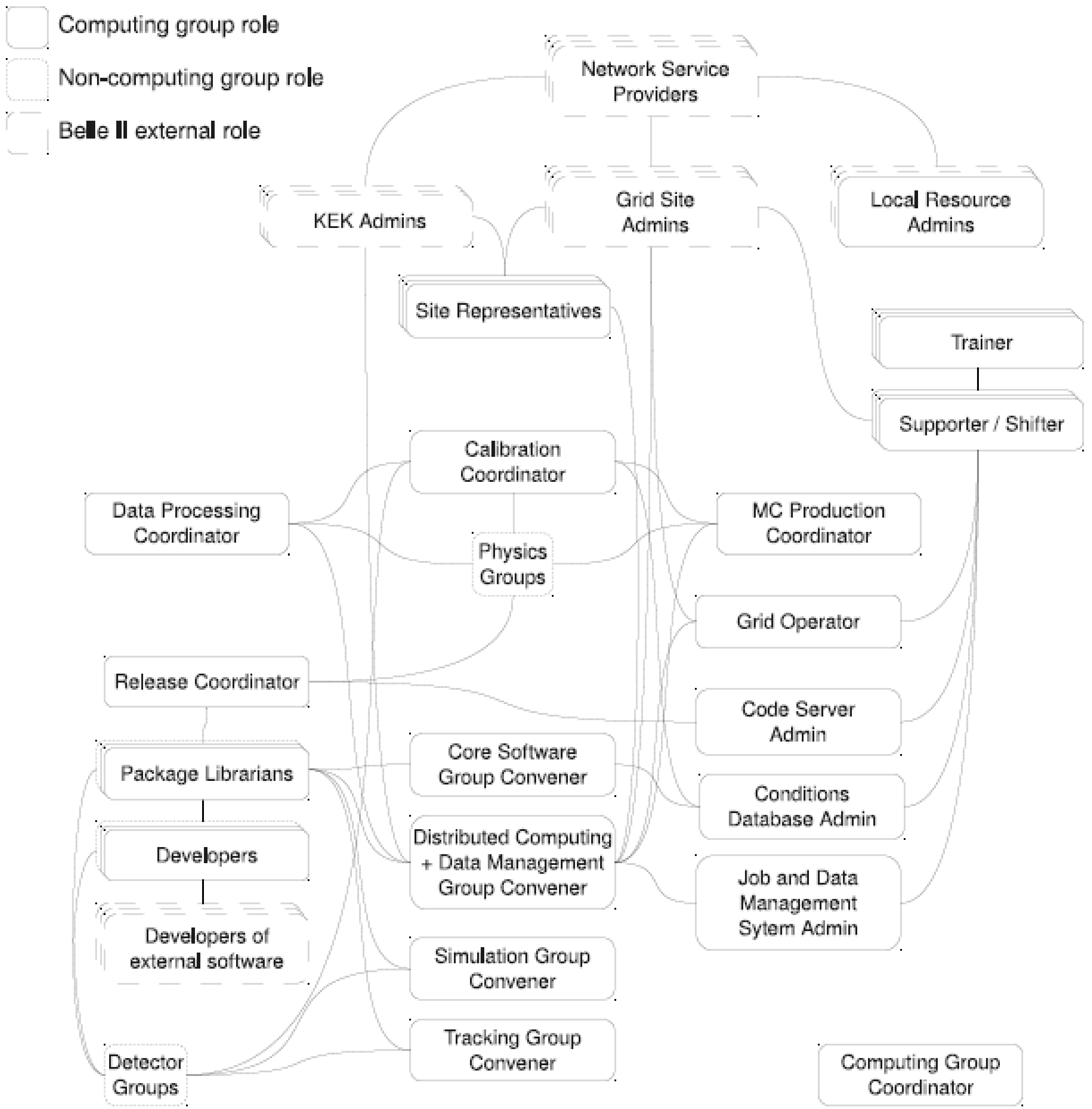
➔ Unique interface to obtain parameters → Gearbox



✔ Same geometry and parameters used for simulation, reconstruction and analysis

➔ Distribution of DB information?

Human Resources



Computing group:

- ~10 in core software team
- ~15 in core distributed computing team (including computing scientists)

Hardware Resources

- Preliminary estimates depend on many unknown parameters (accelerator performance, data reduction, performance of simulation/reconstruction, analysis requirements, ...)